Communication

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 262, No. 28, Issue of October 5, pp. 13377-13380, 1987

Cloning of a Full-length Complementary DNA for an *Artemia salina* Glycine-rich Protein

STRUCTURAL RELATIONSHIP WITH RNA BINDING PROTEINS*

(Received for publication, May 4, 1987)

Marilyn Cruz-Alvarez and Angel Pellicer‡

From the Department of Pathology and Kaplan Cancer Center, New York University Medical Center, New York, New York 10016

Overlapping cDNAs have been isolated containing all the coding sequences for *Artemia salina* protein GRP33, a glycine-rich protein (16.6 mol % glycine), with a molecular weight of 32,992. GRP33 is closely related to HD40, the major protein component of *Artemia* heterogeneous nuclear ribonucleoprotein particles, and shares certain characteristics with other RNA binding proteins. The C-terminal region (123 amino acids) contains 39 glycine residues. This region has multiple arginine residues flanked by glycines, resembling the glycine-dimethylarginine clusters present in other RNA binding proteins. Secondary structure predictions for the protein reveal two distinct domains: a hydrophilic C-terminal domain with an extended conformation and a larger N-terminal domain with a number of α-helices and β-sheets.

In eukaryotic cells, heterogeneous nuclear RNA is associated with a defined set of nuclear proteins to form ribonucleoprotein particles or complexes (hnRNPs),[1] which can be recovered from purified nuclei as substructures with a relatively homogeneous sedimentation coefficient of 30–40 S (1, 2). A major fraction of the proteins from these particles consists of a class of immunologically cross-reactive peptides with molecular weights between 30,000 and 45,000 (3, 4). The amino acid compositions of these proteins are similar, characterized by a high content of glycine (about 20%), very few cysteines, blocked amino termini, and the presence of the modified amino acid dimethylarginine (2, 5–10). hnRNP proteins sharing these characteristics have been found in many divergent species among vertebrates: duck, hamster, mouse, human, as well as in plants, and have been termed "glycine-rich" proteins or "core" hnRNP proteins (6, 10).

Little is known about the specific function of the core proteins; they are thought to be involved in the packaging of heterogeneous nuclear RNA. hnRNP assembly takes place immediately after transcription, and antibodies against these proteins have been shown to inhibit splicing *in vitro* (11, 12) suggesting that core proteins are present in the splicing complexes and are important for RNA processing.

The core proteins in the 30 S particles isolated from several different cell types appear as three major groups (A, B, and C) of closely spaced doublets on SDS-PAGE (2, 13, 14). Two-dimensional gel analysis reveals further complexity, showing some of these proteins contain several differently charged species (6, 15, 16).

The 30 S particles from the brine shrimp *Artemia salina* seem to have a relatively simple protein composition (17). The major protein component has been purified to homogeneity (18). It is a helix-destabilizing protein with a M_r of about 40,000 and has been designated HD40. It binds strongly to single-stranded nucleic acids, forming complexes which are strikingly similar to the native "beads on a string" structures of hnRNPs (18, 19). The biochemical characteristics of HD40: high glycine content, very little cysteine, presence of dimethylarginine, and a blocked amino terminus suggest that this protein is a functional analogue of the hnRNP core proteins from higher eukaryotes (17, 18). Immunoelectrophoresis with a polyclonal antibody raised in rabbits against HD40 reveals the presence of at least three different isoelectric forms of HD40 and three or four other antigenically related proteins (M_r 30,000–40,000) in *Artemia* 30 S particles (17).

As an initial approach for studying hnRNP proteins and their function we had undertaken the cloning of an *A. salina* hnRNP protein. We have previously reported the cloning of a partial cDNA for such a protein using an anti-HD40 antibody (20). We determined the presence of sequences homologous to the cloned cDNA across eukaryotes, from yeast to human by Southern blot analysis, suggesting the conservation of these proteins through evolution.

We describe here the isolation of overlapping cDNAs corresponding to the full-length transcript and the deduced complete amino acid sequence of the encoded protein. Analysis of this sequence and comparison with the only sequence of an hnRNP core protein published so far (21–23) provides some insight into the conserved structural features of these RNA binding proteins.

EXPERIMENTAL PROCEDURES[2]

RESULTS

In order to obtain a full-length cDNA, a new cDNA library was constructed in λ gt11 from *Artemia* total poly(A)$^+$ RNA

* These experiments were supported by National Institutes of Health Grants CA36327 and CA16239. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J03453.

‡ Irma Hirschl/Monique Weil Caulier awardee. To whom correspondence should be addressed.

[1] The abbreviations used are: hnRNP, heterogeneous nuclear ribonucleoprotein; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; BSA, bovine serum albumin; SSC, sodium saline citrate.

[2] Portions of this paper (including "Experimental Procedures" and Figs. 1, 2, and 4) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document No. 87C-203, cite the authors, and include a check or money order for $2.00 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

in addition to the one described previously (20). About 10^4 recombinant phage were screened with a restriction fragment from the previously isolated clone 87HD (20) as a probe. One positive phage, termed λ HD-1, was identified, and this was cloned by plaque purification and two cycles of rescreening. λ HD-1 DNA contained an insert of approximately 1000 base pairs, which was found to overlap 87HD cDNA over 400 base pairs (Fig. 1*A*). Since there were approximately 1000 base pairs that had only been cloned once either in 87HD or λ HD-1, their authenticity was checked by the hybrid selection assay and by Northern blot hybridization in order to discard possible cloning artifacts in these regions. All restriction fragments along the λ HD-1 insert hybrid selected an RNA which was translated *in vitro* into a protein of approximately the same mobility as HD40 on SDS-PAGE. Moreover, the *in vitro* translated protein was immunoprecipitated with anti-HD40 antibodies. In Northern blots the different labeled restriction fragments and clone 87HD detected an identical poly(A)$^+$ RNA species of 1300–1400 nucleotides (data not shown). These results indicate that the entire insert in λ HD-1 represents a cDNA derived from reverse transcription of a single mRNA. When restriction fragments along the cDNA insert in 87HD were analyzed in the same way, the 5′-most restriction fragment failed to hybridize to this RNA. Sequencing of this fragment revealed the presence of a poly(A) tail at its 5′ end, suggesting that it is derived from the adventitious joining of another cDNA in opposite orientation.

Another cDNA library was then constructed in order to obtain a cDNA containing the correct 5′ end. Oligonucleotides 1 and 2 (Fig. 1*A*) were used as probes for the screening of the library. It should be pointed out that the oligonucleotides were contained in a restriction fragment known to hybrid select the correct mRNA and that both of them were shown to be complementary to the same mRNA by Northern blot hybridization (not shown). About 10^4 recombinant phage plaques were screened, and 13 positive phage were identified with both oligonucleotides. Six of these, λ lHD-2 through λ lHD-7, were cloned by plaque purification and rescreening.

The cDNA inserts in λ lHD2 to λ lHD7 were then subcloned into M13mp8 and/or M13mp9 and sequenced by the dideoxynucleotide chain terminator method (31). Five of these cDNAs, those in λ lHD2–lHD6 were found to be identical, while the cDNA in λ lHD7 was a few bases shorter. The overlap of one of the five identical cDNAs with the insert in clone 87HD is shown in Fig. 1*A*. Restriction fragments from the inserts in 87HD and λ HD-1 were also subcloned into M13mp8 and/or M13mp9 and sequenced. The sequencing strategy is shown in Fig. 1*B*.

Primer extension analysis of total *Artemia* poly(A)$^+$ RNA (Fig. 2) shows a single band in a polyacrylamide-urea gel when oligonucleotide 1 was elongated with reverse transcriptase; this band corresponds to the addition of about 132 bases to the primer, the same length of extension seen on the five identical cDNAs isolated (λ lHD2–lHD6). This result indicates that these cDNAs contain the cap site of the RNA.

The entire nucleotide sequence derived from the three overlapping cDNAs is shown in Fig. 3. The presence of a 60-base-long poly(A) stretch and a polyadenylation signal 14 bases upstream indicates that this fragment corresponds to the 3′ end of the mRNA. The sequence in Fig. 3 represents, therefore, that of a full-length cDNA. The length of the mRNA is 1208 bases without the poly(A) tail, in reasonable agreement with the size that had been previously estimated from its mobility in denaturing agarose gels and methyl mercury hydroxide-sucrose gradients (20). The possible open reading frames were analyzed; the longest one contains 939

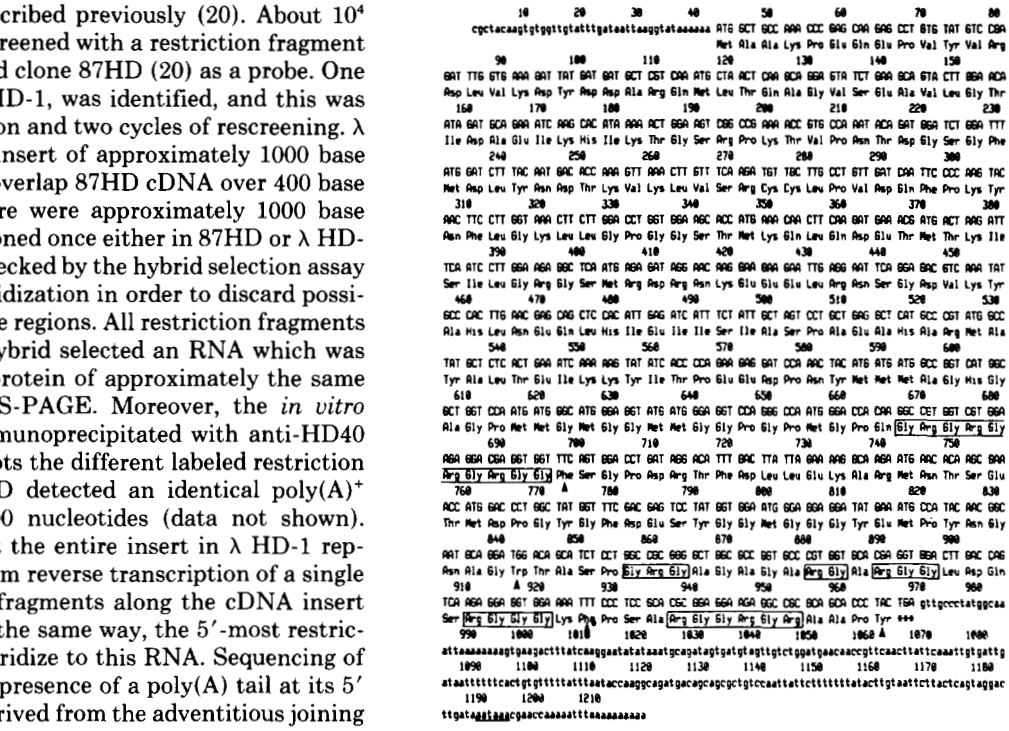

FIG. 3. **Full-length cDNA sequence and amino acid sequence deduced from it.** The whole cDNA sequence is shown starting at the cap site and including 10 A residues of the poly(A) tail. The polyadenylation signal is *underlined*. The sequence corresponding to the 5′- and 3′-untranslated regions of the RNA is indicated by *lower case letters*. Glycine-arginine clusters are shown in *boxes*. Aromatic amino acids situated no more than 6 residues from the glycine-arginine groups are indicated by *arrowheads*.

bases, from base 28 to base 966 as numbered in Fig. 3. The two other reading frames contain a large number of termination codons.

The deduced amino acid sequence starting at the first AUG codon from the 5′ end of the mRNA is also shown in Fig. 3. This first AUG is found in the consensus context (AXXATGG (32)) and has been designated as the initiation site for translation. Two in-frame termination codons (positions 22–24 and 25–27) are found upstream of this initiation codon. There are 924 bases to the first in-phase termination codon corresponding to an open reading frame coding for 308 amino acids. Based on the deduced amino acid composition, the molecular weight of the protein would be 32,992. In view of its high glycine content (16.6%), it has been termed GRP33 (glycine-rich protein, M_r 33,000).

As shown in Fig. 4, the predicted secondary structure (33) for GRP33 reveals two distinct domains. The N-terminal domain contains several possible regions of α-helix and β-sheet; the smaller C-terminal domain would be expected to have an extended conformation. The hydropathy index has been determined along the GRP33 sequence (34) and as shown in Fig. 4, the C-terminal region is essentially hydrophilic.

DISCUSSION

We have previously used a polyclonal antibody raised in rabbits against a major protein component of *A. salina* hnRNPs to screen a cDNA library from the same species. A partial cDNA clone had been identified containing sequences coding for a protein which appeared to be identical to HD40 according to several criteria: the same electrophoretic mobility on SDS-PAGE, common antigenic determinant(s), and virtually identical products of partial proteolysis (20). We have

now obtained overlapping cDNAs containing all the coding sequences. However, the deduced molecular weight of the protein is 33,000 rather than about 40,000. This discrepancy can be explained by the fact that the *in vitro* translated protein shows an anomalously slow mobility on SDS-PAGE and is only distinguishable from the slower migrating HD40 upon electrophoresis on long gels. GRP33 and HD40 are in any case different proteins since several tryptic peptides of HD40 have been recently sequenced[3] and are not present in GRP33. There are several antigenically related proteins in 30 S hnRNP particles of *Artemia* since antibodies raised against purified HD40 recognize on immunoelectrophoresis three or four other proteins of slightly smaller molecular weights than HD40 (17). Western blot analysis of a whole cell extract prepared from *Artemia*-developed cysts also shows the existence of several proteins cross-reacting with the anti-HD40 antibody used for the screening of the cDNA library (data not shown). The relationship between these proteins is not yet known. A similar situation prevails with respect to antigenically and biochemically related groups of hnRNP core proteins present in other species (3, 4, 6, 16). They may represent post-translational modifications of a single gene, alternative splicing products, or products of related genes. GRP33 may then be the precursor of one or more of the *Artemia* proteins which cross-react with anti-HD40 antibodies.

GRP33 shares with a number of hnRNP core proteins what seems to be one of their typical conserved features, a high glycine content, 16.6 mol %, whereas the average frequency of glycine in eukaryotic proteins is 7.6 mol % (PIR Protein Sequence Database). Furthermore, 76.5% of the glycine residues are within the 123-amino acid C-terminal region of the protein. The only complete sequence of a core hnRNP protein published so far, that of rat A1, shows the same unequal distribution of glycine, with 76.9% of the total glycine residues in the 124-amino acid C-terminal region of the molecule (21).

The C-terminal region of GRP33 also shows an unusual content of arginines (10.6%), which are clustered with glycine residues (Fig. 3) and might be methylated *in vivo*. Two other nuclear proteins (rat A1 and nucleolin) show similar Gly-Arg clusters in the C-terminal regions of the molecules (21, 35). Nucleolin, a 110,000 M_r protein which resembles the hnRNP core proteins with respect to the presence of dimethylarginines and a high glycine content, seems to be associated with preribosomal RNA in the nucleus (35). In nucleolin, as in other proteins containing dimethylarginines, *e.g.* a 34,000-dalton nuclear scleroderma antigen from hepatoma cells (36) and human myelin basic protein (37), most of the methylated arginines are surrounded by glycines, suggesting that an adjacent glycine might be required for the methylation of an arginine.

Relatively close to the Gly-Arg groups in this C-terminal domain are several aromatic amino acids: phenylalanine, tryptophan, and tyrosine (Fig. 3). Such aromatic amino acids have been shown to be involved in the binding of some proteins to single-stranded nucleic acids through intercalation of the aromatic residues with the nucleotide bases (38).

The structural characteristics of the C-terminal domain of GRP33, a mostly hydrophilic region with a predicted extended conformation, are shared by two other RNA binding proteins (rat A1 (21) and nucleolin (35)) and are consistent with this region being on the exterior of the molecule and perhaps capable of interacting with nucleic acid.

Recent studies demonstrate that there is a close relationship between eukaryotic single-stranded DNA binding proteins and hnRNP proteins (22, 39). It has been shown that the sequence of the calf thymus single-stranded DNA binding protein, UP1, is identical to that of the 195-amino-acid-long N-terminal domain of hnRNP protein A1 (21–23). Despite the absence of the C-terminal glycine-rich domain, UP1 retains the ability of binding single-stranded nucleic acids, particularly DNA, suggesting that the C-terminal domain may modulate the specificity of the protein to bind RNA over single-stranded DNA.

A similarity search (40) with the National Biochemical Research Foundation Protein Data Base showed the highest homology with an Epstein-Barr virus nuclear antigen (41): 33.6% identity with GRP33 in a 119-amino-acid overlap. Both the N- and C-terminal regions of the Epstein-Barr protein contain repeating Gly-Arg units, and the protein has a high affinity for single-stranded DNA.

In view of the scarcity of sequence information, the complete sequence of this glycine-rich protein which shares structural features with some nuclear RNA binding polypeptides should further the understanding of protein-nucleic acid and protein-protein interactions in hnRNPs.

Acknowledgments—We want to thank R. Lake for his excellent technical assistance, I. Guerrero, J. Leon, and A. Villasante for helpful discussions, and Dr. W. Szer for his continuous help, encouragement, and critical reading of the manuscript.

Note Added in Proof—After the submission of our paper, the complete sequence and primary structure of a human nuclear ribonucleoprotein particle C protein were published by Swanson *et al.* (Swanson, M. S., Nakagawa, T. Y., LeVan, K., and Dreyfuss, G. (1987) *Mol. Cell. Biol.* **7,** 1731–1739).

[3] B. Merril, A. Kumar, W. Szer, and K. R. Williams, unpublished data.

REFERENCES

1. Pederson, T. (1974) *J. Mol. Biol.* **83,** 163–183
2. Beyer, A. L., Christensen, M. E., Walker, B. W., and LeStourgeon, W. M. (1977) *Cell* **11,** 127–138
3. Jones, R. E., Okamura, C. S., and Martin, T. E. (1980) *J. Cell Biol.* **86,** 235–243
4. Leser, G. P., Escara-Wilke, J., and Martin, T. E. (1984) *J. Biol. Chem.* **259,** 1827–1833
5. Karn, J., Vidali, G., Boffa, L. C., and Allfrey, V. G. (1977) *J. Biol. Chem.* **252,** 7307–7322
6. Martin, T. E., Pullman, J. M., and McMullen, M. D. (1980) in *Cell Biology* (Prescott, D. M., and Goldstein, L., eds) Vol. 4, pp. 137–174, Academic Press, Orlando, FL
7. LeStourgeon, W. M., Lothstein, L., Walker, B. W., and Beyer, A. L. (1981) in *Cell Nucleus* (Busch, H., ed) Vol. 9, pp. 49–87, Academic Press, Orlando, FL
8. Martin, T. E., Billings, P. B., Levey, A., Ozarslan, S., Quinlan, T. J., Swift, H. H., and Urbas, L. (1974) *Cold Spring Harbor Symp. Quant. Biol.* **38,** 921–932
9. Economidis, I. V., and Pederson, T. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 1599–1602
10. Raziuddin, Thomas, J. O., and Szer, W. (1982) *Nucleic Acids Res.* **10,** 7777–7789
11. Choi, Y. D., Grabowski, P. J., Sharp, P. A., and Dreyfuss, G. (1986) *Science* **231,** 1534–1539
12. Sierakowska, H., Szer, W., Furdon, P. J., and Kole, R. (1986) *Nucleic Acids Res.* **14,** 5241–5254
13. Dreyfuss, G., Choi, Y. D., and Adam, S. A. (1984) *Mol. Cell. Biol.* **4,** 1104–1114
14. Choi, Y. D., and Dreyfuss, G. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 7471–7475
15. Fuchs, J.-P., Judes, C., and Jacob, M. (1980) *Biochemistry* **19,** 1087–1094
16. Wilk, H. E., Werr, H., Friedrich, D., Kiltz, H. H., and Schafer, K. P. (1985) *Eur. J. Biochem.* **146,** 71–81
17. Thomas, J. O., Raziuddin, Sobota, A., Boublik, M., and Szer, W. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 2888–2892
18. Marvil, D. K., Nowak, L., and Szer, W. (1980) *J. Biol. Chem.* **255,** 6466–6472
19. Nowak, L., Marvil, D. K., Thomas, J. O., Boublik, M., and Szer, W. (1980) *J. Biol. Chem.* **255,** 6473–6478
20. Cruz-Alvarez, M., Szer, W., and Pellicer, A. (1985) *Nucleic Acids Res.* **13,** 3917–3930
21. Cobianchi, F., Sen Gupta, D. N., Zmudzka, B. Z., and Wilson, S. H. (1986) *J. Biol. Chem.* **261,** 3536–3543
22. Kumar, A., Williams, K. R., and Szer, W. (1986) *J. Biol. Chem.* **261,** 11266–11273
23. Riva, S., Morandi, C., Tsoulfas, P., Pandolfo, M., Biamonti, G., Merrill, B., Williams, K. R., Multhaup, G., Beyrenther, K., Werr, B., and Schafer, K. P. (1986) *EMBO J.* **5,** 2267–2273
24. Gubler, U., and Hoffman, B. J. (1983) *Gene (Amst.)* **25,** 263–269
25. Huynh, T. V., Young, R. A., and Davis, R. W. (1984) in *DNA Cloning: A*

Practical Approach (Glover, D. M., ed) pp. 49–78, IRL Press, Wash. D. C.
26. Rigby, P. W., Dieckmann, M., Rhodes, C., and Berg, P. (1977) *J. Mol. Biol.* **113,** 236–251
27. McBride, L. J., and Caruthers, M. H. (1983) *Tetrahedron Lett.* **24,** 245–248
28. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual*, pp. 76–122, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
29. Villasante, A., Wang, D., Dobner, P., Dolph, P., Lewis, S. A., and Cowan, N. J. (1986) *Mol. Cell. Biol.* **6,** 2409–2419
30. Messing, J. (1983) *Methods Enzymol.* **101,** 20–78
31. Sanger, F., Coulsen, A. R., Barrell, B. G., Smith, A. J. H., and Roe, B. (1980) *J. Mol. Biol.* **143,** 161–178
32. Kozak, M. (1986) *Cell* **36,** 993–1005
33. Chou, P. Y., and Fasman, G. D. (1974) *Biochemistry* **13,** 222–244
34. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132
35. Lapeyre, B., Amalric, F., Ghaffari, S. H., Venkatarama Rao, S. V., Dumbar, T. S., and Olson, M. O. J. (1986) *J. Biol. Chem.* **261,** 9167–9173
36. Lischwe, M. A., Ocha, R. L., Reddy, R., Cook, R. G., Yeoman, L. C., Tan, E. M., Reichlin, M., and Busch, H. (1985) *J. Biol. Chem.* **260,** 14304–14310
37. Baldwin, G. S., and Carnegie, P. R. (1971) *Science* **171,** 579–581
38. Chase, J. W., and Williams, K. R. (1986) *Annu. Rev. Biochem.* **55,** 103–136
39. Pandolfo, M., Valentini, O., Biamonti, G., Morandi, C., and Riva, S. (1985) *Nucleic Acids Res.* **13,** 6577–6590
40. Lipman, D. J., and Pearson, W. R. (1985) *Science* **227,** 1435–1441
41. Hennesy, K., and Kieff, E. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 5665–5669

SUPPLEMENTARY MATERIAL TO:

Cloning of a full length cDNA for an Artemia salina glycine-rich protein: structural relationship with RNA binding proteins

by Marilyn Cruz-Alvarez and Angel Pellicer

EXPERIMENTAL PROCEDURES

Construction and screening of cDNA libraries

λgt11 cDNA libraries were prepared as described (24) using poly (A)+ RNA isolated from Artemia salina embryos (20). Screening of the first library was done in standard conditions (25), using as probe a nick-translated (26) ^{32}P-labeled Nru I-Bam HI restriction fragment (specific activity 1-3 x 10^8 cpm/ μg) from the previously isolated clone 87HD (20).

Two oligonucleotides were synthesized (27) with sequences complementary to that of the coding strand, as determined from the previously isolated cDNAs (Fig. 1). Oligonucleotide 1 (a 30-mer) was used as a primer for the construction of the second cDNA library. Oligonucleotides 1 and 2 (a 26-mer) were 5'-labeled with T4 polynucleotide kinase (28) and used as probes for the screening of this library. Filters were blocked for 6 hours at 42 C in 10 x Denhardt's solution (1 x Denhardt's solution: 0.1% Ficoll, 0.1% polyvinyl-pyrrolidone, 0.1% BSA), 20 mM sodium phosphate pH 6.8, 20% formamide and 5 x SSC (1 x SSC: 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0). Each set of duplicate filters was hybridized with both oligonucleotides in the same solution as for the blocking ($5x10^5$-10^6 cpm/ml), for 16-18 hours at 42 C. Filters were washed with 2xSSC, 0.1% SDS, twice at room temperature, twice at 37 C and twice at 47 C.

RNA Primer Extension

Oligonucleotide 1 was ^{32}P-labeled with polynucleotide kinase (28). Approximately 5 ng of labeled oligonucleotide, 10^5 cpm, were mixed with 5 μg of total Artemia poly (A)+ RNA in a reaction (25 μl) containing 50 mM Tris-HCl pH 8.3, 10 mM $MgCl_2$, 150 mM KCl, 10 mM DTT, and incubated at 42 C for 1 hour with 17 units of avian myeloblastosis reverse transcriptase (Life Sciences). The reaction mixture was extracted with phenol, precipitated with ethanol and analyzed directly on a 6% polyacrylamide-8 M urea gel.

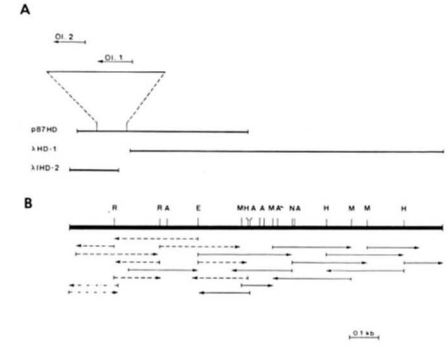

Figure 1. Restriction map and sequencing strategy of p87HD, λHD-1 and λ1HD-2. (A) The overlapping of the cDNA inserts from the positive clones isolated from the three different cDNA libraries is shown. The positions of the oligonucleotides complementary to the coding strand, O1.1 (a 30-mer) and O1.2 (a 26-mer), which were used for the construction and screening of the second cDNA library, are also indicated with respect to the 87HD insert. For simplicity only the relevant portion of this last cDNA is shown. (B) Restriction sites used for subcloning of fragments into M13 and sequencing are shown. Restriction sites are: A, Ava II; E, Eco RI; H, Hae II; M, Mbo II; N, Nco I; R, Rsa I. Continuous lines represent sequencing of the cDNA in lambda HD-1, (- - - -) represents sequencing of the cDNA in clone 87HD, (-.-.-.) represents sequencing of the cDNA in lambda 1HD-2. When a sequence stretch has been determined several times from overlapping fragments, only a few of these fragments are shown.

Restriction mapping and sequencing

Plasmid p87HD and λHD-1 DNAs were purified through CsCl gradients (28). Small scale DNA preparations from phages λ1HD-2 to λ1HD-7 were as described (29). Restriction fragments or the entire cDNAs were subcloned into M13mp8 and/or M13mp9 vectors (30). Single stranded DNAs were prepared from recombinant phage and sequenced by the dideoxynucleotide chain terminator method (31). The universal 17-mer M13 sequencing primer was purchased from New England Biolabs. Sequence data were analyzed using the computer programs of Bionet.

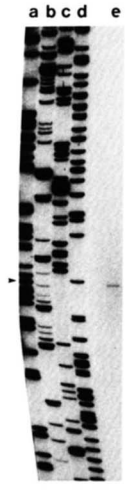

Figure 2. Primer extension analysis. Artemia salina total poly(A)+ RNA was transcribed with reverse transcriptase using ^{32}P-labeled oligonucleotide 1 as primer. The reaction products were analyzed on a 6% polyacrylamide-urea gel (lane e). Lanes a,b,c and d: dideoxy sequence (A,C,G and T respectively) of a clone of known sequence used as size marker. The arrowhead indicates the band corresponding to 162 bases.

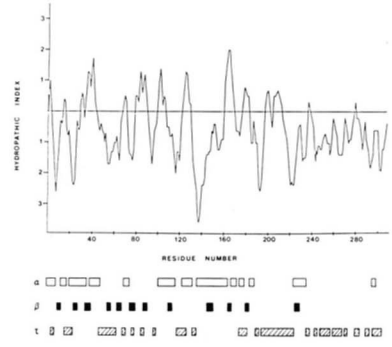

Figure 4. Hydropathy and secondary structure of GRP33. The hydropathy index was plotted according to Kyte and Doolittle (34), with a window size of six residues. The secondary structure is that predicted by the Chou and Fasman algorithm (33).